

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 1, 2003, 08:49:06 ; Search time 93.5 Seconds  
(without alignments)  
6121.913 Million cell updates/sec

Title: US-09-768-781-2

Perfect score: 2543

Sequence: 1 atgaacacaagaccacaaca.....caaggcaagtgctgtctga 1389

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DRV=xlp  
-Q/cgn2\_1/USFTO.spool/US09768781/runat\_01042003\_084730\_4939/app\_query.fasta\_1.1543  
-DB=SPTREMBL\_21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09768781 @CN\_1\_1\_158 @runat\_01042003\_084730\_4939 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_21:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phase:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	953.5	37.5	446	11 Q9QXY7	Q9qxy7 mus musculus

	2	158.5	6.2	395	4	Q9H6D3
	3	141.5	5.5	505	4	Q96P28
	4	140.5	5.5	362	4	Q8TBA0
	5	131	5.2	439	5	O17386
	6	129	5.1	362	11	Q8RI18
	7	121.5	4.8	783	10	Q9SAK8
	8	120.5	4.7	382	17	Q9V2C2
	9	120.5	4.7	745	10	O48539
	10	120.5	4.7	943	10	Q942C7
	11	119.5	4.7	424	8	Q9B8Z8
	12	119	4.7	223	4	Q9NUG5
	13	118.5	4.7	424	8	Q94UX1
	14	116.5	4.6	424	8	Q94UX2
	15	116.5	4.6	424	8	Q94UX6
	16	116.5	4.6	506	16	Q9KCW3
	17	116.5	4.6	2143	10	Q9CBA6
	18	115.5	4.5	411	3	Q9P6N5
	19	115.5	4.5	424	8	Q94UX8
	20	115.5	4.5	498	8	O47543
	21	114	4.5	508	16	Q98QC2
	22	112	4.4	459	8	O21706
	23	111.5	4.4	452	8	Q9B509
	24	111.5	4.4	472	2	O06039
	25	111.5	4.4	785	10	Q9SA37
	26	111	4.4	459	8	O21845
	27	111	4.4	459	8	O21707
	28	111	4.4	459	8	Q955T7
	29	111	4.4	2151	10	Q8RVL2
	30	110.5	4.3	459	8	O21708
	31	110.5	4.3	494	12	Q9QJL1
	32	110	4.3	684	16	Q8XM39
	33	110	4.3	740	16	Q8ZQN3
	34	109.5	4.3	459	8	Q955T8
	35	109.5	4.3	461	16	Q97E21
	36	109.5	4.3	483	5	Q19463
	37	109	4.3	558	12	Q994E7
	38	109	4.3	1145	3	Q9P6W3
	39	108.5	4.3	741	16	Q8X7V5
	40	108.5	4.3	858	16	Q9XKH1
	41	108	4.2	740	16	Q8Z874
	42	108	4.2	1776	5	Q963L7
	43	107.5	4.2	459	8	Q955T9
	44	107.5	4.2	511	2	Q9XDK3
	45	107.5	4.2	572	8	Q8SEY6

## ALIGNMENTS

RESULT 1

Q9QXY7	ID	Q9QXY7	PRELIMINARY;	PRT;	446 AA.
	AC	Q9QXY7;			
	DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
	DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
	DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)			
	DE	KX antigen (1810038K39RIK protein).			
	GN	XXH OR XK OR 1810038K19RIK.			
	OS	Mus musculus (Mouse).			
	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	OX	NCBI_TaxID=10090;			
	RN	[1]			
	RP	SEQUENCE FROM N.A.			
	RC	STRAIN=BALB/C; TISSUE=SKELETAL MUSCLE;			
	RX	MEDLINE=20009522; PubMed=10541802;			
	RA	Collec E., Colin Y., Carbonnet F., Hattab C., Bertrand O.,			
	RA	Carton J.P., Kim C.L.;			
	RT	"Structure and expression of the mouse homologue of the XK gene."			
	RL	Immunogenetics 50:16-21(1999).			
	RN	[2]			
	RP	SEQUENCE FROM N.A.			
	RC	STRAIN=BALB/C; TISSUE=SKELETAL MUSCLE;			
	RC	Le Van Kim C., Collec E., Colin Y.;			

RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.

[3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hata A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaado I., Pesole G., Quackenbush J.,  
 RA Schraml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carrinci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR ENBL; AF155511; AAF14527.1; -;  
 DR ENBL; AK007734; BAB25222.1; -;  
 DR MGD; MGI:103569; Xkh.  
 SQ SEQUENCE 446 AA; 51114 MW; D785PB7B9E28B988 CRC64;

#### Alignment Scores:

Pred. No.: 5,39e-83 Length: 446  
 Score: 953.50 Matches: 181  
 Percent Similarity: 64.72% Conservative: 85  
 Best Local Similarity: 44.04% Mismatches: 134  
 Query Match: 37.50% Indels: 11  
 DB: 11 Gaps: 4

US-09-768-781-2 (1-1389) x Q9QXV7 (1-446)

QY 136 TTTCATTTAGCATCTTTTCCACCTTTTGTACTGTGGGAGGCTGTCATCTGCTTG 195  
 Db 3 PheProAlaSerValIleAlaSerValPheLeuPheValAlaGluThrAlaAlaLeu 22  
 QY 196 TACATGTTAGATCTATCGAAGAATAGTGAACCTTACTGTGATGACATACACCTTTCT 255  
 Db 23 TyrLeuSerSerThrTyrArgSerAlaGlyAspArgMetTrpGlnValLeuLeu 42  
 QY 256 TTCTTTATGTTTTCATCTATTCATGTCAGTGCCTTACCTTATTTTGTCCAGAGATCTA 315  
 Db 43 PheSerLeuMetProCysAlaLeuValGlnPheThrLeuLeuPheValHisArgAspLeu 62  
 QY 316 GCCAAGATAAACCGCTATCATTTATTTATGCTTAATCTCTTGGGACCTGTTTACAGA 375  
 Db 63 SerArgAspArgProLeuAlaLeuLeuMethIleLeuLeuGlnLeuGlyProLeuTyrArg 82  
 QY 376 TGTGTTGAGGCCATGATTAAGTACTCTACATCTGGGAAGAAGAGACGAGAGAGCC 435  
 Db 83 CysCysGluValPheCysIleTyrCys-----GlnSerAspGlnAsnGluGluPro 99  
 QY 436 TATGTCAGCTCACCCGNAAGAAG---ATGCTAATAGTGGGAGGAGGTCTGATAGAA 492  
 Db 100 TyrValSerIleThrLysLysArgGlnMetProLysAspGlyLeuSerGluGluValGlu 119  
 QY 493 TGGAGGTGGGCCACTCCATCCCGACCTGCTATGTCACCGCAATGCTCAAAAGATG 552  
 Db 120 LysGluValGlyGlnAlaGluGlyLysLeuIleThrHisArgSerAlaPheSerArgAla 139  
 QY 553 TCACAGATCCAAAGCCTTCTCGGGTCAAGTGCAGCCAGCTGACCTATCAGCTTATGTAGC 612  
 Db 140 SerValIleGlnAlaPheLeuGlySerAlaProGlnLeuThrLeuGlnLeuTyrIleThr 159  
 QY 613 CTGATCTCTGCAGAGGTTCCCTCGGTAGAGTTGTCTAATGGTATTTCCTCGGTATCT 672

Db 160 ValLeuGluGlnAsnIleThrThrGlyArgCysPheIleMetThrLeuSerLeuSer 179  
 QY 673 GTCACCTATGTTGGGACCCCTTTTGGCAATATGTTGGTATCCAGATCAAGTACGATCTAC 732  
 Db 180 IleValTyrGlyAlaLeuArgCysAsnIleLeuAlaIleLysIleLysTyrAspGluTyr 199  
 QY 733 AAGATTGCGCTTGGCCACTAGAGTCTCTGTCATCACCATCTGGCGACATTTGGAGATC 792  
 Db 200 GluValLysValLysProLeuAlaTyrValCysIlePheLeuTrpArgSerPheGluLeu 219  
 QY 793 ACTTCCCGCTCTCTGATTCTGTTGCTCTTCTTCAGCCACTTTTGAATTTGAAGCTCTGCC 852  
 Db 220 AlaThrArgValIleValLeuValLeuPheThrSerValLeuLysIleTrpValAla 239  
 QY 853 TTCCTAGTGTCAACTCTCTCTGATCATCTCTTTGAGCCCTGGATTAAGTTCTGGAGAGT 912  
 Db 240 ValIleLeuValAsnPhePheSerPhePheLeuTyrProTrpIleValPheTrpCysSer 259  
 QY 913 GGTGCCAGATGCCCAATTAACATTGAGAAAACCTTCAGCCGGTCCGACCTCTGGTGGTC 972  
 Db 260 GlySerProPheProGluAsnIleGluLysAlaLeuSerArgValGlyThrThrIleVal 279  
 QY 973 CTGATTTTCAGTCACCATCTCTATGCTGGCATCAACTTCTCTTGTGCTGCTGCTTTCAG 1032  
 Db 280 LeuCysPheLeuThrLeuLeuTyrAlaGlyIleAsnMetPheCysTrpSerAlaValGln 299  
 QY 1033 TTGAGTTGGCAGACAGAGATCTCTGTCACAAAGGCGAAGACTGGGACATATGCGCTG 1092  
 Db 300 LeuLysIleAspAsnProGluLeuIleSerLysSerGlnAsnTrpTyrArgLeuLeuLeu 319  
 QY 1093 CACTATAGTGTGAGTTGGTAGAAGTGTATCATGCTTCTGTTGTTTAAAGTTCTTGA 1152  
 Db 320 TyrTyrMetThrArgPheIleGluAsnSerValLeuLeuLeuLeuTyrTrpPhePheLys 339  
 QY 1153 GTGAAGTGTACTGAACTACTCTCTGATTCCTTGTGCTGCTGCTGCTTATTCCTTAT 1212  
 Db 340 ThrAspIleTyrMetTyrValCysAlaProLeuLeuIleLeuGlnLeuLeuGlyTyr 359  
 QY 1213 CTGATTTTCATGCTGCTCATGCTCTTTTTCAGTACTTGCATTCATTCGCTGCTCATCT 1272  
 Db 360 CysThrGlyIleLeuPheMetLeuValPheTyrGlnPhePheHisProCysLysLysLeu 379  
 QY 1273 TTCACCATATAGTAGAC-----TACCTCCATTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 Db 380 PheSerSerValSerGluSerPheArgAlaLeuLeuArgCysAlaCysTrpSerSer 399  
 QY 1321 CACCTCGGACGAGGTGAGAACTCAGAGCCA 1353  
 Db 400 -----LeuArgArgLysSerSerGluPro 407  
 RESULT 2  
 Q9H6D3 PRELIMINARY; PRT; 395 AA.  
 AC Q9H6D3;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE CDNA: FLJ22371 fis, clone HRC06680 (Hypothetical 44.7 kDa protein)  
 DE (Similar to hypothetical protein FLJ10307).  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC TISSUE=SKIN;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SKIN;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK026024; BAB15326.1; -;  
DR EMBL; BC013379; AAH13379.1; -;  
DR EMBL; BC028564; AAH28564.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 395 AA; 44654 MW; 23199BAEEA6964C6 CRC64;

Alignment Scores:  
Pred. No.: 1.87e-06 Length: 395  
Score: 158.50 Matches: 92  
Percent Similarity: 38.33% Conservativity: 82  
Best Local Similarity: 20.26% Mismatches: 161  
Query Match: 6.23% Indels: 119  
DB: 4 Gaps: 21

US-09-768-781-2 (1-1389) x Q9H6D3 (1-395)

QY 91 TCTCTGAGAGATGTCATCCGTGGAGCAACCCCGATTACTTTTCCATTAGCATC 150  
Db 8 AlaLeuLeuArgAspLeuValLeuGly----- 16

QY 151 CTTTCTCCACCTTTTGTACTGTGGGAGGCGTCATCTGCTTTGTACATGGTTAGATC 210  
Db 17 ValLeuGlyThrAlaAlaPheLeuLeuAspLeuGlyThrAspLeuTrpAlaAlaValGln 36

QY 211 TATCGAAAGATAGTGAACCTTACTGTGATGACATACACCTTTTCTTTCTTTATGTTCA 270  
Db 37 TyrAlaLeuGlyArgTyrLeuTrpAlaAlaLeuValLeuAlaLeuLeuGlyLeuAla 56

QY 271 TCCATTATGTCAGTTG---ACCCTCATTTTGTCCACAGAGAT-----CTAGCC 318  
Db 57 SerValAlaLeuGlnLeuPheSerTrpLeuTrpLeuArgAlaAspProAlaGlyLeuHis 76

QY 319 AAAGATAAACCG-----CTATCATTTATTTATGATCTAATCTCTCTTGGACCT 366  
Db 77 GlySerGlnProProArgArgCysLeuAlaLeu---LeuHisLeuLeuGlnLeuGlyTyr 95

QY 367 GTTATCAGATGTTTGGAGGCCATGATTAAGTACCTCACACTGTGGAGAAAGAGGAGCAG 426  
Db 96 LeuTyrArgCysValGlnLeuLeuArgGlnGlyLeuValTrp-----GlnGln 112

QY 427 GAGGAGCCCTATGTACAGCCTCACCCGAAAGAGATGCTAATAGATGGCGAGGAGTGTG 486  
Db 113 GluGluPro----- 115

QY 487 ATAGATGGGAGTGGGCGCACTCCATCCGACCCCTGGTATGACCCGCAATGCTTACAA 546  
Db 116 SerGluPheAspLeuAlaTyrAla---AspPheLeuAlaLeuAsp-----IleSer 131

QY 547 CGTATGTCACAGATCCAAGCCTCTCTGGGCTAGTCCCGCCAGCTGACCTATCAGCTCTAT 606  
Db 132 MetLeuArgLeuPheGluThrPheLeuGluThrAlaProGlnLeuThrLeuValLeuAla 151

QY 607 GTGAGCCTGATCTCTGACAGAGTTCCCTGGGTAGAGTTGTGTAATGGTATTTTCCCTG 666  
Db 152 IleMetLeuGlnSerGlyArgAlaGluTyrTyrGlnTrpValGlyIleCysThrSerPhe 171

QY 667 GTATCTGTACCTATGGGCGCCACCTTTTGCATATATGTTGGCTATCCAGATCAAGTACGAT 726  
Db 172 LeuGlyIleSerTrpAlaLeuLeu----- 179

QY 727 GACTAC-----AAGATTCGCTTGGGCGCACTAGAA 756  
Db 180 AspTyrHisArgAlaLeuArgThrCysLeuProSerLysProLeuLeuGlyLeuGlySer 199

QY 757 GTCTCTGCATCACCATCTGGCGGAGTGGAGATCACTTCCCGCCTCTGATTCTGTG 816

Db 200 SerValIleTyrPheLeuTrpAsnLeuLeuLeuLeuTrpProArgValLeuAlaValAla 219  
QY 817 CTCTTCTCAGCCACTTTG---AAATTGAAGCTGTGCCCTTCCTAGTGTCAACTTCTCTG 873  
Db 220 LeuPheSerAlaLeuPheProSerTyrValAlaLeuHisPheLeuGlyLeuTrpLeuVal 239  
QY 874 ATCATCTCTTTAGCCCTGGATTAAAGTTCTGGAGAAAGTGTGCCAG---ATGCCCAAT 930  
Db 240 LeuLeuLeu-----TrpVal-----TrpLeuGlnGlyThrAspPheMetProasp 254  
QY 931 AACATTGAGAAAACTTCAGCCGGTGGCAGCTCTGGTGGTCTGATTTCAGTCCACCATC 990  
Db 255 ProSerSerGluTrpLeuTyrArgValThrVal-----AlaThrIle 268  
QY 991 CTCTATCTGGCATCAACTTCTCTGTGTGTGTCAGCTTTCGAGTTGCGAGTTGGCAGACAGA 1050  
Db 269 LeuTyr-----PheSerTrpPheAsn----- 275  
QY 1051 GATCTCGTCGACAAAGGGCAGAACTGGGGACATATGGCCCTGCACATATAGTGTGAGGTTG 1110  
Db 276 -----ValAlaGluGlyArgThrArgGlyArgAlaIleIleHisPheAlaPheLeuLeu 293  
QY 1111 GTAGAGAAATGTATCATGTCTTGTGTTTAAAGTTCTTTTGGAGTGAAGTGTACTGAAT 1170  
Db 294 SerAspSerIleLeuLeuValAlaThrTrpValThrHisSer-----Ser 308  
QY 1171 TACTGTATCTCTTGTATGTCCTTGCCTGCAGCTCATTAAT-----GCTTATCTG 1215  
Db 309 TrpLeuProSerGlyIleProLeuGlnLeuTrpLeuProValGlyCysGlyCysPhePhe 328  
QY 1216 ATTTCCATTTGGCTTCATGTCTCTCTTCTCCAGTACTTGTGATCCATTCGCGTCACTCTTC 1275  
Db 329 LeuGlyLeuAlaLeuArgLeuValTyrTyrHisTrpLeuHisPro----- 343  
QY 1276 ACCATAATGTAGTACTACCTCCATTGTGTCTGTCTGCCAGCACCCCTCGGACCCAGG 1335  
Db 344 -----SerCysCysTrpLysProAspProAspGln 353  
QY 1336 GTTGAGAACTCAGAGCCACCCCTTTGAGACTGAAGCAAGGCAA 1377  
Db 354 ValAspGlyAlaArgSerLeuLeuSerProGluGlyTyrGln 367

RESULT 3  
Q96P28 ID Q96P28 PRELIMINARY; PRT; 505 AA.  
AC Q96P28;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1889 protein (Fragment).  
GN KIAA1889.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=21456161; PubMed=11572484;  
RA Nagase T., Kikuno R., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XXI.  
RT The complete sequences of 60 new cDNA clones from brain which code for  
RT large proteins.";  
RL DNA Res. 8:179-187(2001).  
DR EMBL; AB067476; BAB67782.1; -;  
FT NON\_TER 1  
SQ SEQUENCE 505 AA; 56911 MW; D764B96BC29B3623 CRC64;

Alignment Scores:  
Pred. No.: 8.48e-05 Length: 505  
Score: 141.50 Matches: 65  
Percent Similarity: 42.41% Conservativity: 69

```
Best Local Similarity: 20.57%
Query Match: 5.56%
DB: 4
Mismatches: 121
Indels: 61
Gaps: 13
```

US-09-768-781-2 (1-1389) x Q96P28 (1-505)

	QY	331	CTATCATTTATTGCATCTAATCCTCTTGGGACCTGTGTATTCAGATGTTTGAGGCCCATG	390
	Db	110	LeuGlnSerLeuIleHisIIIELEuGlnLeuGlyGlnIleTrpArgTyPheHisThrIle	129
	QY	391	ATTAAGTACTCTCACACTGTGGAGAAGAGGACGACGAGGAGCCCTATGTACGCTCACC	450
	Db	130	TyrLeuGlyIleArgSerArgGlnSerGlyGlu	140
	QY	451	CGAAAGAAGATGCTAATAGATGCGGAGGAGGTCTGATAGAATGGGAGGTGGGCCACCTCC	510
	Db	141	AsnAspArgTrpArgPheTyTrpLysMetValTyrglu	153
	QY	511	ATCCGGACCCCTGGCTATGCAACCGCATGCCACAAAGTATGTCACAGATCCAAGCCTTC	570
	Db	154	TyrAlaAspValSerMet	166
	QY	571	CTGGGCTCAGTGGCCCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGTT	630
	Db	167	LeuGlusErAlaProGlnLeuValLeuGlnLeuCysIlellevalGlnThrHisSerIeu	186
	QY	631	CCCTGGGTAGAGTTGTGCTAAATGGTATTTTCCTCGTATCTGTACCTATGGGCCCACC	690
	Db	187	GlnAlaLeuGlnGlyPheThrAlaAlaAlaSerLeuValSerLeuAlatrpp--AlaLeu	205
	QY	691	CTTTGCCAATATGTTGGCTATCCAGATCMAGTAGCATCTACNAGATTCGCCTTGGGCCA	750
	Db	206	AlaSerTyrglnLysAlaLeuArgAspSerArgAspLyslys-----Pro	221
	QY	751	CTAGAAGTCTCTGCATCACCATC-----TGGCGGACATTGGAGATCACTTCCCGC	801
	Db	222	IleSerTyrrMetAlaValllelGlnPheCyeTrpHisPhePheThrIleAlaAarg	241
	QY	802	CTCTGATTCCTGGTCTCTCTCAGCCACTTTGAATATGAAGCTGTGCCCTTCTCTAGTG	861
	Db	242	ValIleThrPheAlaLeuPheAlaSerValPheGlnLeuTyrrPheGlyIlePheIleVal	261
	QY	862	CTCAACTTCTGTATCATCCCTTTTGAGCCCTGGATTAAAGTTCTGGAGAAAGTGGTCCCAC	921
	Db	262	LeuHisTrpCysIleMetThrPhe-----TrpIleVal-----	272
	QY	922	ATGCCCAATAAACATTGAGAAAACTTC-----AGCCGGGTGGCACAATGCGCTGCTG	975
	Db	273	HisCysGluThrGluPheCysIleThrLysTrpGluIleValPheAsp	289
	QY	976	ATTTCAGTCACCATCTCTATGCTGGCATCACTCTCTTGTGTGGTCACTTTGAGTTG	1039
	Db	290	MetValValGlyIleIleTyrlle-----PheSerTrpPheAsn-----	302
	QY	1036	AGGTGGCAGACAGATCTCTCGCAAAAGGCAGAACTGGGGACATATGGCTGCGCAC	1099
	Db	303	ValLysGluGlyArgThrArgCysArgLeuPheIleTyrr	315
	QY	1096	TATAGTGTGAGTGGTPAGAGAATGTGATCATGCTCTGGTTTTTAAAGTTCTTTGGAGTG	1155
	Db	316	TyrPheValIleLeuLeuGluAsnThrAlaLeuSerAlaLeuTyrrLeuTyrrLysAla	335
	QY	1156	AAGTGTACTGAATTACTGTCTATTCCTTGATTTGCTTCCAGTACTTTCATCCCA	1260
	Db	336	ProGlnIleAlaAsp---AlaPheAlaIleProAlaLeuCysValValPheSerSerpHe	354
	QY	1213	CTGATTTCCATTCAGTCTCTCTTTTCTCCAGTACTTTCATCCCA	1260
	Db	355	LeuThrGlyValValPheMetLeuMetTyrrTyrAlaPhePheHiPro	370
	RESULT 4			
	QB8BA0			
	ID	QB8BA0	PRELIMINARY;	PRT; 362 AA.

RESULT 4  
Q8TBA0  
ID Q8TBA0

Db	207	AlaGlyLeuAlaMetMetLeuLeuTyr	TyrGlyValLeuHisPro	-----221
Qy	1276	ACCCATAATGTAGTAGACTACCTCCATGGTGCTCGCTGT	1314	
Dd	222	ThrGlyProArgAlaLysIleLeuAlaSerSerCysCys	234	
RESULT 5				
O17386				
ID	O17386	PRELIMINARY;	PRT; 439 AA.	
AC	O17386;			
DT	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	Hypothetical 51.1 kDa protein.			
GN	F08F1.5.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OX	Rhabditidae; Peloderinae; Caenorhabditis.			
OC	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RA	MEDLINE=99069613; PubMed=9851916;			
RX	None;			
RT	"Genome sequence of the nematode C. elegans: a platform for			
RT	investigating biology. The C. elegans Sequencing Consortium."			
RL	Science 282:2012-2018 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RA	Fulton B.;			
RT	"The sequence of C. elegans cosmid F08F1.";			
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RA	Waterston R.;			
RT	"Direct Submission.";			
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF026213; AAB71305.1; -.			
KW	Hypochemical protein.			
SQ	SEQUENCE 439 AA; 51150 MW; E42AC8F17CD2877E CRC64;			
Alignment Scores:				
Pred. No.:	0.000849	Length:	439	
Score:	131.00	Matches:	67	
Percent Similarity:	37.39%	Conservative:	62	
Best Local Similarity:	19.42%	Mismatches:	98	
Query Match:	5.15%	Indels:	118	
DB:	5	Gaps:	15	
US-09-768-781-2 (1-1389) x O17386 (1-439)				
Qy	328	CCGCTATCATTTATTTCATCTAATCCTC-----TTGGGACCTGTTATCAGATGTTTG	381	
Dd	126	ProLenSerLysMetIleValLeuCysIleCysGlnMetGlyProLeuPheTrpTyrTyr	145	
Qy	382	GAGGCCATGATTAAAGTACCTCACACTGTGGAGAAGAAGAGAGCAGGAGGAGGCCTATGTC	441	
Dd	146	LysAlaLeuTyr---TyrGlyTrpMetPheArgLysSerSerAsnGluAsn-----	161	
Qy	442	AGCCTCACCCGAAGAAGATGCTAATAGATGGCGAGAGTGCTGATCAATGGGAGGTG	501	
Dd	162	-----ThraspGlyGluLys-----	166	
Qy	502	GGCCACTCCATCCGGACCCTGGCTATGTCACCCCAAATGCCTTACAAACGTATGTCACAG---	558	
Dd	167	-----ArgLysCysPheSerLysMetValGluAla 176		
Qy	559	-----ATCCAAGCTTCTCGGGCTCACTGCCCGAGCTG	591	
Dd	177	GluArgAspAlaThrLeuLeuArgPhePheGluAlaPheLeuGluSerAlaProGlnLeu	196	

QY 591 ----- 591

Db 197 IleileGlnGlySerileAlaAlaSerTyrPheGlnAsnTyrTyrGlnThrGlyThrTyr 216

QY 592 ACCTATCAGCTCTCATGTGAGCCCTGATCTCTGAGAGGTTCCCTGGGTAGAGTTGTGCTA 651

Db 217 ProTyrTrpLeuTyrPheGlnAlaAlaSer ----- 226

QY 652 ATGTGATTTTCCCTGGGTATCTGTCACCTATGAGGGCCACCCCTTTGCAATATGTGGCTATC 711

Db 227 LeuLeuLeuSerileileSerileSerTyrSerValValValGlnAsnArgSerLeuArg 246

QY 712 CAGATCAAGTACGATGACTACAAGATCCCTTTGGGCCACATAGAGTCCTCTGTCATCAC 771

Db 247 MetileArg ----- AspAspLysValAsnIleTyrProHisGluAlaValLeuGlnPhe 264

QY 772 ATCTGGGGACATTGGAGATCACTCCCGCCCTCCGATTCCTGGTGCTCTCTCTCAGCCACT 831

Db 265 CysTrpArgPheLeuThrIleLeuAlaArgIleIleThrLeuValAla ----- 280

QY 832 TTGAATTTGAAGCTGTGGCCCTCTCTAGTGTCTCAACTCTCTCATCATCTCTTTTGAGGCC 891

Db 281 ----- LeuValLeuIlePhe ----- 285

QY 892 TGGATTAAAGTTCTGGAGAAAGTGGTCCCGACAGATGCCCAATAACATTTGAGAAAACCTTCAGC 951

Db 286 ----- GlyIleAsnValAlaIleHisIleAspAla --- CysThr 297

QY 952 CGGGTCGGCACTCTGGTGTCTGATTTTCAGTCACCATCTCTATGCTGGCATCACTTC 1011

Db 298 HisIleGlnLysLeuLeuLeuIleAsn --- ThrPheIleHisIlePheIleProPhe 316

QY 1012 TCITGCTGGTCAGCTTTGCAGTTTGAGTTGGCAGACAGATCTCGTCACAAAAGGGCAG 1071

Db 317 ----- AsnMetValGluGlyAsnThr 323

QY 1072 AACTGGGACATATGCGCTCGCACTATGATGTGAGGTTGGTAGAGAAATGTCATCATGCTC 1131

Db 324 ArgTrpArgTyrLeu --- ThrAlaTyrSerValGluPheIleGlu ----- MetMet 339

QY 1132 TTGGTTTTTAAGTCTTTGGAGTGAAGTGTTACTGAATTACTGTCATTCTCTGATTGCC 1191

Db 340 LeuValCysTrpLeuLeuProLeuSerLeuAsnThrPheProTyrIleGluLysValGln 359

QY 1192 TTGCAGCTCATATTCTTATCTGATTTTCATTTGGCTTCATGCTCTCTTTCTCCAGTAC 1251

Db 360 ValGlyValProIleSerPheIleAlaGlyIleAlaIleMetMetMetTyrTyrGlnPhe 379

QY 1252 TTGCATCCATTGCGC 1266

Db 380 PheHisProAsnArg 384

RESULT 6

Q8R118

ID Q8R118 PRELIMINARY; PRT; 362 AA.

AC Q8R118;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Similar to KIAA1889 protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;

RP [1]

RN SEQUENCE FROM N.A.

RC TISSUE=COLON;

RA Strausberg R.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC024502; AAH24502.1; -

SQ SEQUENCE 362 AA; 41643 MW; 51D0F86C5E548017 CRC64;

Alignment Scores:

**Alignment Scores:**

```

Pred. No.: 0.00128 Length: 362
Score: 129.00 Matches: 58
Percent Similarity: 40.98% Conservative: 42
Best Local Similarity: 23.77% Mismatches: 100
Query Match: 5.07% Indels: 44
DB: 11 Gaps: 9

US-09-768-781-2 (1-1389) x Q8RI18 (1-362)
QY 559 ATCAAGCTTCTCGGCTCAGTCCAGCTGACCTATCAGCTCTATGTGAGCTGATC 618
DB 14 LeuGluThrPheLeuGluSerAlaProGlnLeuValLeuGlnLeuCysIleMetIleGln 33
QY 619 TCTCAGAGGTTCCCTCGGTGAGATTGTGCTTAATGGTATTTCCTCGTATCTCTCACC 678
DB 34 LysAsnSerAlaGluThrLeuProCysValSerSerValThrSerLeuMetSerLeuAla 53
QY 679 TATGGGGCCACCTTTGCAATATGTTGCTATCCAGATCAAGTACGATGATCAAGATT 738
DB 54 TrpValLeuAlaSerTyrHisLeuLeuLeuArgAspSerArgAspAspLysLys--- 71
QY 739 CGCTTGGGGCCTAGAGTCTCTGTCATCACCATCTCGGGGACATTGGAGATCACTTCC 798
DB 72 SerMetSerTyrArgGlyAlaLeuIleHisLeuPheTrpArgLeuPheThrIleSerSer 91
QY 799 CGCTCTCTGATCTGCTCTCTCAGCCACTTTGAAATGAGGCTGTGCCCTTCCTA 858
DB 92 ArgValIleSerPheAlaLeuPheAlaSerIlePheGlnLeuTyrPheGlyIlePheVal 111
QY 859 GTGCTCAACTCTCGATCATCTCTTTGAGCCCTGATTAAGTTCTGAGAGAGTGGTCC 918
DB 112 ValValHisTrpCysAlaMetAlaPhe-----TrpIleIleHisGlyThrAspPhe 129
QY 919 CAGATGCCCAATACATTGAGAAAACCTTACGCGGTCGGCCTCTCGTGTCTGATT 978
DB 130 CysMetSerLysTrpGluGluIleLeuPheAsnMetVal----- 142
QY 979 TCAGTCAACCTCTATGCTGGCATCAACTTCTCTCTGTCAGCTTTGCGAGTTGAGG 1038
DB 143 ---ValGlyIleValTyr-----IlePheCysTrpPheAsn----- 153
QY 1039 TTGGCAGACAGATCTGTCGACAAAGGGGAGAACTGGGACATATATGGCTGCACTAT 1098
DB 154 -----ValLysGluGlyArgThrArgTyrArgMetPheAlaTyrTyr 167
QY 1099 AGTGTGAGTTGTAGAAATGATCATGTCTTGGTTTTTAAGTTC----- 1146
DB 168 ThrIleValLeuThrGluAsnAlaAlaLeuThrPheLeuTrpTyrPheTyrArgAsnPro 187
QY 1147 -----TTGGAGTGAAGTGTACTGAATTACTCTCATCTCTCTGATTGCC 1191
DB 188 GluSerThrAspSerTyrAlaValProAlaLeuCys-----Cys----- 200
QY 1192 TTGCAGCTCATATTGCTATCTGATTTCCTATGCTTTCATGCTCTCTTCTTCCAGTAC 1251
DB 201 -----ValPheValSerPheValAlaGlyIleThrLeuMetLeuTyrTyrGlyVal 218
QY 1252 TTGCATCATTTG 1263
DB 219 LeuHisProMet 222

RESULT 7
Q9SAK8 PRELIMINARY; PRT; 783 AA.
AC Q9SAK8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T8K14.18 protein.
GN T8K14.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Kremenetskaia I., Luros J., Lee J.M., Gonzalez A., Altafi H.,
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federpiele N.A., Theologis A.,
RT "Arabidopsis thaliana chromosome 1 BAC T8K14 sequence."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007202; AAD30236.1; -.
DR InterPro; IPR000676; Nah_Exchange.
DR Pfam; PF00999; Na_H_Exchange.
SQ SEQUENCE 783 AA; 88203 MW; 0A6846910C381648 CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.00771 Length: 783
Score: 121.50 Matches: 89
Percent Similarity: 34.58% Conservative: 68
Best Local Similarity: 19.60% Mismatches: 166
Query Match: 4.78% Indels: 131
DB: 10 Gaps: 19

```

## US-09-768-781-2 (1-1389) x Q9SAK8 (1-783)

```

QY 256 TTCCTTTATGTTTTCATCATATGTCAGTTGACCTCATTTTGTTCACAGATCTA 315
DB 16 PheAsnProLeuAsnThrMetPheIleGlnMetAlaCysIleLeuValPheSerGlnLeu 35
QY 316 -----GCCAAAGATAAACCGCTATCATTTATTTATGATCTA 351
DB 36 PheTyrLeuLeuLeuLysProCysGlyGlnAlaGlyProValAlaGlnIleLeuAlaGly 55
QY 352 ATCTCTTGTGGACCTGTTATCAGATGTTTGGAGGCCATGATTAACTTCTCACACTGTGG 411
DB 56 IleValLeuSerProValLeuLeuSerArgIleProLysValLysGluPheLeuGln 75
QY 412 AAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 465
DB 76 LysAsnAlaAlaAspTyrTyrSerPhePheSerPheAlaLeuArgThrSerPheMetPhe 95
QY 466 ATAGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 525
DB 96 LeuIleGlyLeuGluValAspLeuHisPhe----- 105
QY 526 ATGCACGCGCAATGCTACAAACGATGATGTCACAGATC-----CAAGCCTTCTCG----- 573
DB 106 MetArgArgAsn---PheLysLysAlaAlaValIleThrLeuSerSerPheValValSer 124
QY 574 GGCTCAGTGGCCCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGGTTCCC 633
DB 125 GlyLeuLeuSerPheAlaSerLeuMetLeuPheIleProLeuPheGlyIleLysGluAsp 144
QY 634 CTGGGTAGAGTTGCTTAATGCTATTTTCCCTGCTATCTGTACCTATGCGGCGCCACCTT 693
DB 145 TyrPheThrPhePheLeuValLeuValThrLeuSerAsnThrAlaSerProValVal 164
QY 694 TGCAATATGTTGGCTATCCAGATCAAGTACGATGATGATCAACAAGATT-----CGCCTT 744
DB 165 ValArgSerIleAla-----AspTrpLysLeuAsnThrCysGluIle 178
QY 745 GGGCCACTAGAGGATCTCTGTCATCACCATCTGGGGGACATTTGGAGATCATCTTCCGCGCTC 804
DB 179 GlyArgLeuThrIleSerCysAlaLeuPheIleGluLeuThrAsnValValLeuTyrThr 198
QY 805 CTGATTTCTGTGCTCTCTCTCAGCCACTTTGAAATGAGGCTGTGCCCTTCTTCTAGTCTC 864
DB 199 IleIleMetAlaPheIleSerGlyThrIleIleLeuGluLeuPheLeuPheLeuAla 218
QY 865 AACTTCTCTGATCATCTCTTGTAG-----CCCTGGATTAACTTCTGGAGAGT 912

```









```
Db 72 ArgProLeuSerTyrLysGlyAlaValAlaGlnValLeuTrpHisLeuPheSerIleAla 91
Qy 796 TCCGCGCTCCATGATCTGGTCTCTTCAGCGCATTTGAAATTAAGAGCTGTGCCCTTC 855
Db 92 AlaArgGlyLeuAlaPheAlaLeuPheAlaSerValTyrLysLeuTyrPheGlyIlePhe 111
Qy 856 CTAGTGTCTCAACTCTCTGATCATCTCTTTGAGCCCTGGATTAAGTTCTGGAGAAGTGT 915
Db 112 IleValAlaHisTrpCysValMetThrPhe-----TrpVal----- 123
Qy 916 GCCAGATGCCCAATACATTGAGAAAACCTTACCGGGTGGCACTCTGGTGTCTCTG 975
Db 124 -----IleGlnGlyGluThrAspPheCysMetSerLysTyrGluGluIleIleTyrAsn 141
Qy 976 ATTTCACTCACCATCTCTATGCTGGCATCACTCTCTGCTGTGTGCTGAGCTTGCAGTTG 1035
Db 142 MetValValGlyIleIleTyr-----IlePheCysTrpPheAsn----- 154
Qy 1036 AGTTGGCAGACAGATCTCTGTCACAAAGGGCAGAACTGGGGACATATGGGCTGCAC 1095
Db 155 -----ValysGluGlyArgSerArgArgMetThrLeuTyr 167
Qy 1096 TATAGTGTGAGTGTGTAGAGAATGTGATGCTGTGGTCTTTTAAGTTCTTTGGAGTG 1155
Db 168 HisCysIleValLeuLeuGluAla-----AlaLeuThrGlyPheTrpTyrSerSer 185
Qy 1156 AAGTGTCTAGTAATCTATGCTATCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1212
Db 186 ArgAsnPheSerThrAspPheTyrSerLeuIleMetValCysValAlaAlaSerSerPhe 205
Qy 1213 CTGATTTCCATGCTTCCATGCTCTCTTTCTTCCAGTACTTGCATCCA 1260
Db 206 AlaLeuGlyIlePhePheMetCysValTyrTyrCysLeuLeuHisPro 221

RESULT 13
Q94UX1 PRELIMINARY; PRT; 424 AA.
AC Q94UX1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE NADH dehydrogenase subunit 4.
GN NAD4
OS Schistosoma japonicum (Blood fluke).
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20349913; PubMed=10899225;
RA Le T.H., Blair D., Agateuma T., Humair P.F., Campbell N.J.,
RA Iwagami M., Littlewood D.T., Peacock B., Johnston D.A., Bartley J.,
RA Rollinson D., Herniou E.A., Zaranga D.S., McManus D.P.;
RT "Phylogenies inferred from mitochondrial gene orders-a cautionary tale
from the parasitic flatworms."
RL Mol. Biol. Evol. 17:1123-1125(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Le T.H., Blair D., McManus D.P.;
RT "Revisiting limited genetic variation within Schistosoma japonicum
populations."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL: AF412213; AAL12155.1;
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam: PF00361; Oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 424 AA; 47725 MW; 1BC880B001235E31 CRC64;
```

Alignment Scores:

Pred. No.: 0.0135 Length: 424

```
Score: 118.50 Matches: 91
Percent Similarity: 36.36% Conservative: 77
Best Local Similarity: 19.70% Mismatches: 155
Query Match: 4.68% Indels: 139
DB: 8 Gaps: 20
US-09-768-781-2 (1-1389) x Q94UX1 (1-424)
Qy 133 ACTTTTCCATTTAGCATCTTTTCTCCACCTTTTGTACTGTGGGAGGCTGCATCTGCT 192
Db 18 ThrPheIlePheLeuLeuIlePheTyrSerSerIleTrpValSerAspSerMetVal 37
Qy 193 TTGTACATGGTTAGTAATCTATCGAAAGAAATAGTAGAACTTACTGGATGACATACACCTTT 252
Db 38 MetValGlyValLysTyrTyrLeuCysAspGlyLeuValIleAspThr----- 54
Qy 253 TCTTCTTTATGTTTTCATCCATATGATGCTCAGTTGACC---CTCATTTTGTCCACAGA 309
Db 55 -----LeuSerCysLeuMetIlePheLeuThrSerIleIleTrpLeuValLeu 70
Qy 310 GATCTAGCCAAAGATAAACCGCTATCATTTATTCATCTAATCTCTTGGACCTGTT 369
Db 71 TrpLeuValGlySerLysAspIleValLeuPhe----- 81
Qy 370 ATCAGATGTTTGGAGGCCATGATTAAGTAC-----CTCACACTGTGGAAG 414
Db 82 IleSerValPheSerAlaMetLeuThrTyrValValSerAsnSerLeuValPheTrpPhe 101
Qy 415 AAAGAGGAGCAGGAGGCCCTATGTCAGCTTACCCGAAAGAGATGCTAATAGATGCG 474
Db 102 PheTyrGluLeuSer-----IleIleSerAlaLeuTyrMetLeuIleValGly 117
Qy 475 -----GAGGAGTGTGTAGATGAGAGTGGGAGTGGCCACTCCATCGGACCTG 522
Db 118 SerProTyrProGluArgTyrIleSerSerTrpTyrPheGlyTyrIle----- 134
Qy 523 GCTATGCACCGCAATGCTACAAAGTATGTACAGATCCAAAGCCTTCTCGGCTCAGTG 582
Db 135 -----LeuLeuSerSerVal 139
Qy 583 CCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGGTTCCTCGGTGATA 642
Db 140 ProLeuLeuLeuGlyIleCysPheIleGlyLeuAsnSerGlySerPhe-----Asn 156
Qy 643 GTTGTGCTAATGTTATTTTCCCTGGTATCTGTCACTATGAGGGCCACCTTTGCAATATG 702
Db 157 ValIleLeuTyrAspLysGlyAspMetCysAspSerTyrGlyAlaPheLeuLeuIle 176
Qy 703 TTG----- 705
Db 177 ValMetPheLeuThrLysIleProValPheProPheHisGlyTrpLeuProLeuValHis 196
Qy 706 -----GCTATCCAGATCAAGTACGATGACTACAAGATTCGCTTGGGCCA 750
Db 197 AlaGluAlaSerSerProValSerIleIleLeuSerGlyTyrIleMetLysLeuGlyLeu 216
Qy 751 CTAGAAGTCTCTTCATCACCATCTGCGGAGCATTTGGAGATCATTCCCGCTCTCGATT 810
Db 217 ValGlyLeuValArgLeuCysGlyTrpLeuLeuIleAsp-----Tyr 230
Qy 811 CTGGTGTCTTCTCAGCCACTTTGAAATTAAGGCTGTGCCCTTCTAGTGTCAACTTC 870
Db 231 IleTyrTyrPheSerThrPheLeuLeuCysTyrSerValValPheLeuVal----- 247
Qy 871 CTGATCATCTCTTTGAGCCCTGGATTAAGTTCTCGAAGAGTGGTGGCCAGATGCCCAAT 930
Db 248 ---AlaAlaValPheGluCysAspSerLysArgTrpLeuAlaTyrLeuSerLeuSerHis 266
Qy 931 AACATTGAGAAAAAATTCAGCGGCTCGGCACTCTGGTGGTCTGCTG----- 975
Db 267 IleLeu-----IleGlyValCysIleLeuLeuThrSerThrTyrCys 280
Qy 976 -----ATTTCAGTCACCATCTCTAT 996
```

```

Db 281 GlyAspTyrLeuAlaPheIleTyrCysLeuGlyHisGlyLeuSerValAlaLeuPhe 300
QY 997 GCTGGCATCAACTCTCTGCTGGTGCAGCTTGGAGTTGGCAGACAGATCTC 1056
Db 301 MetIleIle-----TrpPheGlyTyrGlu----- 308
QY 1057 GTCGACAAAGGGCAGAACTGGGACATATGGCCCTGCACATAGTGTGAGGTTGGTAGAG 1116
Db 309 IleserGlySerArgAsnTrpGlyLeuValIlePheGlyGlyGlyLeuMet 328
QY 1117 AATGTGATCATGCTCTGGTGTCTT-----AAGTTC 1146
Db 329 HisPheIleMetGlyPheValPheLeuAsnValCysGlyPheProAlaLeuGlnPhe 348
QY 1147 TTGGGA---GTGAAAGTGTACTGAATTAC-----TGTCACTCTTCATTGCTTCGAG 1197
Db 349 PheGlyLeuLeuTrpLeuValIleAsnTyrIleThrLeuGlyAspIleIleSerLeu 368
QY 1198 CTCATTATTGCTTATCTGATTTC-----ATTGGCTTCATGCTCTTCTTCAG 1248
Db 369 LeuValSerIleTyrIlePheSerGlySerIleIleGlyPheIleIleTyrGlyLeuVal 388
QY 1249 TACTGTGATCATGCTGCTCCTCTTACCATAAT-----GTAGTAGACTACTTCAT 1302
Db 389 IleCysSerProIleAsnThrSerTyrGlyLeuAspAsnPheLeuPhe 408
QY 1303 TGTGTC 1308
Db 409 CysIle 410

RESULT 14
O94UX2 PRELIMINARY; PRT; 424 AA.
AC O94UX2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE NADH dehydrogenase subunit 4.
GN NAD4.
OS Schistosoma japonicum (Blood fluke).
OC Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2034913; PubMed=10889225;
RA Le T.H., Blair D., Agatsuma T., Humair P.F., Campbell N.J.,
RA Iwagami M., Littlewood D.T., Peacock B., Johnston D.A., Bartley J.,
RA Rollinson D., Herniou E.A., Zariwaga D.S., McManus D.P.,
RT "Phylogenies inferred from mitochondrial gene orders-a cautionary tale
RT from the parasitic flatworms."
RL Mol. Biol. Evol. 17:1123-1125(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Le T.H., Blair D., McManus D.P.;
RT "Revisiting limited genetic variation within Schistosoma japonicum
RT populations."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL; AF121211; AAL12147.1;
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 424 AA; 47669 MW; D197589A496177B8 CRC64;

```

```

Alignment Scores:
Pred. No.: 0.021 Length: 424
Score: 116.50 Matches: 90
Percent Similarity: 36.36% Conservative: 78
Best Local Similarity: 19.48% Mismatches: 155
Query Match: 139 Indels: 139

```

```

DB: 8 Gaps: 20
US-09-768-781-2 (1-1389) x Q94UX2 (1-424)
QY 133 ACITTTCCATTAGCATCCCTTTTCTCCACCTTTTGTACTGTGGGAGGCTGCATCTGCT 192
Db 18 ThrPheIlePheLeuLeuIlePheTyrSerIleIleTrpValSerAspSerMetVal 37
QY 193 TTGTACATGTTAGAACTATCGAAAGAATAGTCAAACCTTACTGGATGACATACACCTTT 252
Db 38 MetValGlyValIleTyrTyrLeuCysAspGlyLeuValIleIleAspThr----- 54
QY 253 TCTTTCTTATGTTTTCATCATTCATGTCAGTTGACC-----CTCATTTTGTCCACAGA 309
Db 55 -----LeuSerCysLeuMetIlePheLeuThrSerIleIleTrpLeuValLeu 70
QY 310 GATCTAGCCAAAGATAAACCGCTATCATTTATTTATGATCTAATCTCTTGGGACCTGTT 369
Db 71 TrpLeuValGlySerIleAspIleValLeuPhe----- 81
QY 370 ATCAGATGTTTGGAGGCGCATGATTAAAGTAC-----CTCACACTGTGGAAG 414
Db 82 IleserValPheSerAlaMetLeuThrTyrValValSerAsnSerLeuValPheTrpPhe 101
QY 415 AAAGAGGAGCAGGAGGAGCCCTATGTCAGCCTCACCCGAAAGAAGATGCTAATAGATGCG 474
Db 102 PheTyrGluLeuSer-----IleIleSerAlaLeuTyrMetLeuIleValGly 117
QY 475 -----GAGGAGTGTGATAGATGAGAGGTGGGCGCCACTCCATCGGACCCCTG 522
Db 118 SerProTyrProGluArgTyrIleSerSerTyrPheGlyTyrIle----- 134
QY 523 GCTATGCACCCCAATGCTACAAACGTATGTCACAGATCCAAAGCCTTCTCGGCTCAGTG 582
Db 135 -----LeuLeuSerSerVal 139
QY 583 CCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGGTTCCCTTGGGTAGA 642
Db 140 ProLeuLeuGlyIleCysPheIleGlyLeuAsnSerGlySerPhe-----Asn 156
QY 643 GTTGTGTAATGGTATTTTCCCTGGTATCTGTCACCTATGGGGCCACCCTTTGCAATATG 702
Db 157 ValIleLeuTrpAspLysGlyAspMetCysAspSerTyrGlyAlaPheLeuLeuIle 176
QY 703 TTG----- 705
Db 177 ValMetPheLeuThrLysIleProValPheProPheHisGlyTrpLeuProLeuValHis 196
QY 706 -----GCTATCCAGATCAAGTACGATGATGATGATGATGATGATGATGATGATGATGAT 750
Db 197 AlaGluAlaSerSerProValSerIleIleLeuSerGlyTyrIleMetLysLeuGlyLeu 216
QY 751 CTAGAAGTCTCTGCATCACCATCTGGCGGACATTTGGAGATCAGTTCGGGCTCTCTGATT 810
Db 217 ValGlyLeuValArgLeuCysGlyTrpLeuLeuIleAsp-----Tyr 230
QY 811 CTGTGTCTCTCTCAGCCACTTTGAAATTGAAGCTGTGCCCTTCTAGTGTCTCACTTC 870
Db 231 IleTyrTyrPheSerThrPheLeuLeuCysTyrSerValValTyrLeuVal----- 247
QY 871 CTGATCATCTCTTGTAGCCCTCGGATTAAGTTCTGGAGAAGTGGTGGCCAGATGCCAAT 930
Db 248 ---AlaAlaValPheGluCysAspSerLysArgTrpLeuAlaTyrLeuSerLeuSerHis 266
QY 931 AACATTGAGAAAACTTCCAGCGGGTGGCAGCTCTGTGGTGGTCTCTG----- 975
Db 267 IleLeu-----IleGlyValCysIleLeuLeuThrSerThrTyrCys 280
QY 976 -----ATTTCAGTCCCATCTCTAT 996
Db 281 GlyAspTyrLeuAlaPheIleTyrCysLeuGlyHisGlyLeuSerValAlaLeuPhe 300
QY 997 GCTGGCATCAACTCTCTGCTGGTGCAGCTTGGAGTTGGCAGACAGATCTC 1056

```



Db	309	IleSerGlySerArgAsnTrpGlyIleLeuValLysIlePheGlyGlyLeuIleMet	328
Qy	1117	AAATGTGATCATGCTCTGGTGT---AAGTTC	1146
Db	329	HisPheIleMetGlyPheValPheLeuAsnValCysGlyPheProProAlaLeuGlnPhe	348
Qy	1147	TTTGGG---GTGAAAGTGTACTGAATAC-----TGTCAATCTCTGATGGCTTGCAG	1197
Db	349	PheGlyGluLeuTrpLeuValIleAsnTrpIleThrLeuGlyAspIleIleSerLeuLeu	368
Qy	1198	CTCATATTGCTTATCTGATTCC-----ATTGGCTTCATGCTCCCTTTCTTCCAG	1248
Db	369	LeuValserIleTyrllePheSerGlySerIleIleGlyPheIleIleTyrglyLeuVal	388
Qy	1249	TACTTGCATCATTTGGCGCTCATCTTCCACCCATAAT---GTATGATAGACTACTCCAT	1302
Db	389	IleCysSerProIleAsnThrSerTyrglutyrSerGlyGlyLeuAspAsnPheLeuPhe	408
Qy	1303	TGTGTC	1308
Db	409	CysIle	410

Search completed: April 1, 2003, 08:57:34  
Job time : 108.5 secs